

Introduction to LREC 2018 by Nicoletta Calzolari
Chair of the 11th edition of LREC
ELRA Honorary President

Welcome to the 11th edition of LREC in Miyazaki, first LREC in Asia!

LREC 20th Anniversary

It is the LREC 20th Anniversary and LREC has become one of the most successful conferences of the field. Data are pervasive in Natural Language Processing and Language Technology: we call our data Language Resources (LR). But when LREC was started by ELRA, in 1998 in Granada, from an idea of Antonio Zampolli and Joseph Mariani, it was really a new adventure and a challenge. There were well established big conferences but he thought that the new emerging field of Language Resources deserved its own dedicated forum. In the keynote talk I gave at LREC1998 I could say: “the infrastructural role of Language Resources as the necessary common platform on which new technologies and applications can be based is nowadays widely recognised.” This could not have been said only few years before. I had the pleasure and the honour of being involved in LREC from the beginning, first as member of the Program Committee and since 2004 as Conference Chair.

LREC is probably the most influential ELRA achievement, and a service with the major impact on our community. Also through LREC, ELRA contributes to shape our field, making the Language Resource field a scientific field in its own right.

Why LREC in Asia this time? AFNLP (the Asian Federation of NLP) asked us if we could hold an LREC in Asia as the best instrument to promote Language Resources in Asia. We were glad to accept this challenge and here we are.

Some LREC2018 figures

As expected given the change in continent, we did not break any record this time, but the figures are not far from the previous. We received 1102 submissions for the main conference, 34 workshop proposals and 8 tutorial proposals.

A very large part of our community was involved in the reviewing effort, to be able to assign few papers per reviewer: 1263 colleagues accepted to act as reviewers (more than in 2016) out of 1796 invited (268 declined and 265 unfortunately not answering). Few reviewers did not complete the task (only 26 reviews missing, not so bad), but knowing that this always happens we recruited some pinch-reviewers able to act at the last moment: a good move to keep for the future.

The Program Committee has also been enlarged with 3 colleagues from Japan and one from USA. We had as usual a very hard job, examining about 3300 reviews, to understand – beyond the scores and in particular when they greatly differed – the relevance, the novelty, but also the appropriateness for an oral or poster presentation. I am sure we made mistakes, every reviewing effort is not immune from subjectivity, but as usual we discussed in a face to face meeting not only general policies, criteria and how to be consistent, but also borderline cases to arrive at agreed decisions. Overall we all believe we received in average good submissions. We have in the main program 718 papers: 188 Orals and 530 Posters.

We also have 29 Workshops and 5 Tutorials.

I am proud that around 1100 participants have already registered at the end of April, similar to last time. They come from 63 countries. The Japanese are the largest group and in general there is a larger participation from Asian countries, in particular China, as we obviously hoped.

These figures have a clear significance. The field of Language Resources and Evaluation is very alive and constantly flourishing.

LREC acceptance rate: a motivated choice for an inclusive conference

The LREC acceptance rate, 65% this year, is different from other major conferences but for us it is a motivated decision. This is one of the reasons why LREC succeeds to provide a comprehensive picture of the field and to show how it is evolving. For us it is important not only to hear about new methodologies but also to understand how various methods or resources are able to spread, for which purposes, usages, applications and for which languages. Multilingualism – and equal treatment of all languages – is an essential feature of LREC, as it is the attempt of putting the text, speech and multimodal communities together as well as academics and industrials. LREC wants to be an “inclusive” conference.

Quality is not undermined by our acceptance rate: in 2017 Google Scholar Metrics h5-index, LREC ranks 4th in Computational Linguistics top conferences (5th considering ArXiv which is the first).

LREC2018 Novelties

Industry Track

Because of the interest in joining forces between academy and industry, this time we decided to experiment with a new Industry Track. We spoke about this at last LREC with Linne Ha from Google and we asked her if she wanted to organise it for LREC2018.

Special Speech Session

A special session on “Speech resources collection in real-world situations” was proposed to us by Kikuo Maekawa and Yuichi Ishimoto (National Institute for Japanese Language and Linguistics): we gladly accepted also to strengthen the participation of the speech community at LREC.

Oriental-COCOSDA Conference

Also O-COCOSDA is organised together with LREC. We spoke with Satoshi Nakamura, its chair, at last LREC and he kindly offered to organise it jointly with LREC. We are very pleased of this also because it is another opportunity to reach the Asian speech community.

ELRA Individual Members Assembly

ELRA has recently introduced “individual membership” in addition to institutional membership. This was decided to give a voice inside ELRA to the large LREC community and offer them its services. The first assembly of ELRA individual members is held on the first day of the conference.

The LREC Club

From the answers received, it seems that the LREC Club of those who attended all editions, the really faithful ones, is composed of 23 members. I want to thank them for their loyalty!

LREC2018 Trends

I quickly sketch here, as I always do, my perception – subjective and impressionistic – of LREC2018 trends and how certain topics fluctuate from an LREC to the other. The comparison with previous years shows the topics with steady progress, or even great leaps forward, the stable ones and those more affected by the fashion of the moment.

Trends in LREC2018 topics

Among the areas that continue to be trendy and are even increasing I can mention:

- Less-Resourced Languages
- Social Media analysis, appearing in 2012 and since then constantly growing
- Semantics in general and in particular Sentiment, Emotion and Subjectivity
- Information extraction, Knowledge discovery, Text mining are booming
- Lexicons (in its various forms)
- Discourse, Dialogue, Conversational systems and Interactivity
- Multimodality, also for Less-Resourced languages
- Tools, Systems, Applications for various purposes: Question Answering, Summarisation, etc.
- Evaluation methodologies
- Computer Aided Language Learning

Stable “usual” topics, some very well-represented, others in the medium/low range, are:

- Infrastructural issues, policies, strategies and Large projects: topics that receive special attention at LREC, differently from other major conferences
- Corpus creation, annotation, use, ...
- Speech related topics, a little increasing but not as much as we would like
- Sign language (also a very successful workshop)
- Crowdsourcing
- Anaphora and Coreference
- Temporal and Spatial annotation

New trends for this LREC:

- Digital Humanities (new for LREC in 2016, now increased)
- Bibliometrics, Scientometrics, Infometrics
- Language Modelling

Decreasing topics with respect to the past, even if some still numerous:

- Grammar and syntax and also Treebanks that had a big increase in 2016
- Multilinguality and Machine Translation, very high in 2016
- Ontologies
- Standards and metadata are much less represented
- Linked data, a new topic in 2014, seems no longer so fashionable
- Web services and workflows also no longer so popular

The recognition given by the LR community to infrastructural issues, strategies and policies may be also due to the fact that we must often work in large groups, for many languages, we must build on each other work, connect various resources and tools, make available what already exists and use standardised formats. Infrastructures (on many dimensions) are really needed for our field to progress: to pay proper attention to these issues is another distinguishing feature of LREC.

LREC-related initiatives

Proceedings in Thomson Citation Index

Since 2010 the LREC Proceedings are included in CPCI (Thomson Reuters Conference Proceedings Citation Index): an important achievement, providing a better recognition to all LREC authors and useful in particular for young colleagues.

LRE Journal and LREC

After each LREC we ask to the authors of papers suggested by the 3 reviewers as appropriate for LRE if they want to submit an extended version to the *LRE journal*, coedited by Nancy Ide and myself. I am glad to report that also the journal has a large and increasing number of submissions, testifying the great interest for the field of LRs and Evaluation.

Citation of Language Resources

Also this year we encouraged citations of LRs in a special References section (introduced in 2016), providing recommendations on how to cite. I hope this becomes normal practice, to keep track of the relevance of LRs but also to provide due recognition to those working on LRs.

LRE Map and Share your Language Resources

As usual we encouraged descriptions of LRs in the LRE Map, an innovative instrument introduced at LREC2010 with the aim of monitoring the wealth of data and technologies developed and used in our field. And we ask, since 2014, to share the LRs with all the community.

In this LREC about 1000 LRs have been described in the Map. Just few hints at some data in the 2018 Map: WordNets, Wikipedia, Prague TreeBank are the most cited LRs; Corpora are by far the most frequent type (half of the LRs); and about 85% of the LRs are in some way available (not bad).

Replicability of research results

I believe that research is strongly affected also by infrastructural (meta-research) activities as those mentioned above. With these initiatives I hope we are able to promote in our field what is in use in more mature disciplines, i.e. ensure proper documentation and reproducibility of research results as a normal practice. ELRA and LREC are thus influential in strengthening the Language Resources and Evaluation scientific ecosystem and fostering sustainability.

Acknowledgments

As usual, it is my pleasure to express here my deepest gratitude to all those who made this LREC2018 possible and hopefully successful.

I first thank the Programme Committee members, not only for their dedication in the huge task of selecting the papers, the workshops and tutorials, but also for the constant involvement in the various aspects around LREC. I wish to thank each of them, the new ones: Chris Cieri, Koiti Hasida, Hitoshi Isahara, Take Tokunaga, and obviously the “old” ones: Khalid Choukri, Thierry Declerck, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, Stelios Piperidis. A particular thanks goes to Jan Odijk, who has been so helpful in the preparation of the program.

I thank ELRA and the ELRA Board: LREC is a major service from ELRA to all the community!

A very special thanks goes to Sara Goggi and H el ene Mazo, chairs of the Editorial Committee, for the dedication and competence in managing the so many tasks they have in their hands, and the capability to tackle the many big and small problems of such a large conference (not an easy task). They are the two pillars of LREC, without whose commitment for many months LREC would not happen. So much of LREC organisation is on their shoulders, and this is visible to all participants.

I am especially grateful to Hitoshi Isahara and Kyoko Kanzaki, the Local Committee, for their great efforts in dealing with so many local matters and for their patience and true commitment, looking at so many details. We owe them a lot in organising a successful LREC.

My appreciation goes also to the distinguished members of the Advisory Board, chaired by Makoto Nagao, for their support and precious advices.

I am very grateful to the Local Liaison Committee, representing a number of Asian associations and organisations that have supported LREC, in particular with dissemination tasks.

I express my great gratitude to the Sponsorship Committee, and to all the Sponsors that have helped with financial support, believing in the importance of our conference.

I thank the Japanese Ministry of Education, Culture, Sports, Science and Technology for its precious support.

I am particularly grateful to the local authorities, the Governor of Miyazaki prefecture and the Mayor of Miyazaki city, very supportive of LREC since my first visit in 2016. We met them, and other local authorities, several times. We thank them also for their financial support.

In my many visits to Miyazaki I have been impressed by the great sense of hospitality of locals, and among them I wish to thank at least Manmatsu Hayashi and Rie Saita for their great help and kindness.

Also on behalf of the Program Committee, I praise our impressively large Scientific Committee. They did a wonderful job.

I thank the workshop and tutorial organisers, who complement LREC of so many interesting events.

I thank the organisers of the Industry Track, of the Special Speech session and of O-COCOSDA.

A big thanks goes to all the LREC authors, who provide the “substance” to LREC, and give us such a broad picture of the field.

This time I really want to thank also Softconf (Rich Gerber and Paolo Gai) and their constant efforts to make START a better tool for us. I greatly appreciated the new feature for plagiarism detection: I must say it was useful to detect some paper too similar to others ...

I thank the European Commission for the interest in our conference, and hope that funding agencies will be impressed by the quality and quantity of initiatives in our sector that LREC displays, and by the fact that the field attracts the best groups of R&D from all continents. The success of LREC for us means the success of the field of Language Resources and Evaluation.

I finally thank the two institutions that always dedicate a great effort to LREC: ELDA in Paris and ILC-CNR in Pisa. Without their commitment LREC would not be possible. The last, but not least, thanks are thus, in addition to H el ene Mazo and Sara Goggi, to all the others who – with different roles – have helped and will help during the conference: Roberto Bartolini, Damien Bihel, Irene De Felice, Riccardo Del Gratta, Pawel Kamocki, Val erie Mapelli, Monica Monachini, Vincenzo Parrinelli, Vladimir Popescu, Valeria Quochi, Caroline Rannaud, Alexandre Sicard.

And lastly, my final words are for all the LREC2018 participants, the true protagonist of LREC. Now LREC is in your hands. I hope that you discover new paths, that you perceive the vitality and strength of the field, that you have fruitful conversations (conferences are useful also for this) and most of all that you profit of so many contacts to organise new exciting work and projects in the field of Language Resources and Evaluation ... which you will show at the next LREC.

This Japanese LREC in Miyazaki has a sort of Mediterranean flavour, typical of LREC. I am sure you will appreciate the Japanese great hospitality and kindness. And I hope that Miyazaki will enjoy the invasion of LRECCers!

With all the Programme Committee, I welcome you at LREC2018 and wish you a very fruitful Conference.

Enjoy LREC2018 in Miyazaki!